TELL ME SOME MORE: EXPLORING HOW VERBAL ABILITY AND ITEM VERIFIABILITY INFLUENCE RESPONSES TO BIODATA QUESTIONS IN A HIGH-STAKES SELECTION CONTEXT

JULIA LEVASHINA
Kent State University

FREDERICK P. MORGESON
Michigan State University

MICHAEL A. CAMPION
Purdue University

Although self-rated or self-scored selection measures are commonly used in selection contexts, they are potentially susceptible to applicant response distortion or faking. The response elaboration technique (RET), which requires job applicants to provide supporting information to justify their responses, has been identified as a potential way to minimize applicant response distortion. In a large-scale, high-stakes selection context ($N = 16,304$), we investigate the extent to which RET affects responding on a biodata test as well as the underlying reasons for any potential effect. We find that asking job applicants to elaborate their responses leads to overall lower scores on a biodata test. Item verifiability affects the extent to which RET decreases faking, which we suggest is due to increased accountability. In addition, verbal ability was more strongly related to biodata item scores when items require elaboration, although the effect of verbal ability was small. The implications of these findings for reducing faking in personnel selection are delineated.

There are a variety of different methods that could be used to collect information about job applicants in order to make selection decisions. Some measures (e.g., cognitive ability measures, job knowledge tests) have objectively correct answers and ask applicants to demonstrate what is being measured. Other measures (e.g., job interviews, assessment centers) have answers judged subjectively by third parties (such as interviewers or assessors) where applicants are asked to either describe or demonstrate what is being measured. Another class of measures (e.g., personality tests, biodata measures) has answers rated subjectively by the applicants (i.e., self-ratings) where applicants are asked to self-assess, self-rate, or self-score on what is being measured.

Correspondence and requests for reprints should be addressed to Julia Levashina, College of Business Administration, Kent State University, Kent, OH 44242-0001; jlevashi@kent.edu.

Self-rated or self-scored selection measures have become increasingly prevalent in the last several decades. In addition to demonstrating positive criterion-related validity (Bobko, Roth, & Potosky, 1999; Hunter & Hunter, 1984; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997), these kinds of selection tests have addressed some of the limitations associated with more cognitively oriented measures (e.g., adverse impact; Reilly & Chao, 1982; Roth, Bevier, Bobko, Switzer, & Tyler, 2001; Schmidt, 1988). Yet, despite these positive attributes, self-rated measures have other limitations. In particular, they are potentially susceptible to response distortion or faking (McFarland & Ryan, 2000; Morgeson et al., 2007; Rosse, Stecher, Miller, & Levin, 1998). Self-rated measures are susceptible to being faked by motivated job applicants because the right answer may be apparent to the applicants. In contrast, motivated job applicants cannot fake the right answer on measures with objectively correct answers.

Response distortion or faking can attenuate the criterion-related validity of tests (Peeters & Lievens, 2005), negatively affect construct validity (Stark, Chernyshenko, Chan, Lee, & Drasgow, 2001), and affect who is hired (Rosse et al., 1998). As such, scholars have sought to control job applicants' tendencies to inflate their responses through both reactive and proactive approaches (Lievens & Peeters, 2008; Vasilopoulos, Cucina, & McElreath, 2005).

The reactive approach includes methods designed to detect the occurrence of faking. This includes social desirability scales (Crowne & Marlowe, 1960), impression management scales (Paulhus, 1998), and bogus items (Anderson, Warner, & Spencer, 1984; Levashina, Morgeson, & Campion, 2009). In all these approaches, items designed to measure faking are embedded in selection tests, and after job applicants complete the tests, their scores are corrected in order to remove social desirability or impression management biases.

The proactive approach includes methods that seek to prevent faking by either decreasing the motivation to fake (e.g., using warnings; Dwight & Donovan, 2003) or by decreasing the opportunity to fake such as by imposing forced-choice response formats (Christiansen, Burns, & Montgomery, 2005; Travers, 1951) or using nontransparent test items (Alliger & Dwight, 2000; Seeman, 1952). Although some of the proactive methods date to the 1950s, the proactive approach has only recently received renewed attention from researchers. This attention is primarily due to the growing body of knowledge that the most commonly used reactive methods to control faking may be ineffective and can themselves be faked (Ellingson, Sackett, & Hough, 1999; Schmitt & Oswald, 2006).

A recently proposed proactive method of reducing score inflation on biodata measures involves requiring respondents to elaborate their responses (Schmitt & Kunce, 2002; Schmitt et al., 2003). This response

elaboration technique (RET) requires job applicants to provide supporting information to justify their answers to test questions (Schmitt & Kunce, 2002). To date, four empirical studies have been conducted to test the usefulness of this technique. Three of these studies investigated its impact on biodata responses (Ramsay, Schmitt, Oswald, Kim, & Gillespie, 2006; Schmitt & Kunce, 2002; Schmitt et al., 2003), and one study investigated its impact on situational judgment test responses (Lievens & Peeters, 2008). The reported results provide initial empirical support for the usefulness of RET and show that requiring elaboration for biodata items reduces test scores. Although the results are promising, several issues remain to be addressed.

First, we do not know how elaboration affects responses of job applicants in a high-stakes selection environment where applicants are naturally motivated to enhance their scores on selection tests in order to increase the likelihood of being hired. Previous studies of the elaboration technique have not used participants who were completing biodata measures under the real expectations that their responses would contribute to selection decisions. Schmitt and Kunce (2002) recruited their participants via ads in the local media and paid them to complete the testing. Schmitt et al. (2003), Ramsay et al. (2006), and Lievens and Peeters (2008) asked undergraduate students to participate in their studies. It is possible that the experimental context could have affected the reported results. For example, the decreased scores on the biodata items with elaborations could be due to the participants' intentional response deflation (as hypothesized), or it could be due to participants' lack of motivation to engage in demanding item responding and to write the required descriptions for elaborated items given the fact that there were few meaningful outcomes at stake (like job offers).

Job applicants, on the other hand, are likely to be highly motivated and willing to exert the effort to respond to elaboration items. They are more likely to use any opportunity to enhance and inflate their scores on selection tests in order to increase the likelihood of being hired (Rosse et al., 1998). Therefore, it is important to study the RET in a high-stakes selection environment. Past research has acknowledged this, calling for further investigation into the underlying processes of the RET and the usefulness of this technique in actual applicant settings (Lievens & Peeters, 2008; Schmitt et al., 2003).

Second, deflated scores on elaborated items may be due to increased demands on applicant verbal abilities. There is some empirical evidence suggesting that proactive approaches to control response distortion, such as using warnings and forced-choice response formats, increase the complexity of response decisions, resulting in increased relationships between test scores and cognitive ability (Vasilopoulos et al., 2005; Vasilopoulos,

Cucina, Dyomina, & Morewitz, 2006). Asking job applicants to elaborate their responses can make it more difficult for them to respond to the test items, thereby enhancing the cognitive loading of the biodata test and result in the increased relationship between biodata and verbal ability scores. But because biodata measures are typically designed to be less cognitively oriented, their usefulness may be inadvertently affected when job applicants are asked to elaborate their responses. Therefore, it is important to examine the relationship between the RET and demands on verbal abilities.

Finally, across past studies the use of the RET has been confounded with the subset of items chosen for elaboration. As Schmitt et al. (2003, p. 985) acknowledge, "The items that were chosen for elaboration were more objective and verifiable." The same potential limitation was noted by Ramsay et al. (2006, p. 284): "required elaboration may have been on items that were more verifiable." It is important to further investigate the confounding of elaboration and item verifiability, in part because verifiable (e.g., "What was your grade point average in college?") and nonverifiable (e.g., "How often have you looked for a new way to complete an assignment?") items may provide different opportunities for job applicants to engage in response distortion. As such, it is not clear if lower mean scores on biodata items found in past research were due to the required elaboration or item verifiability. Acknowledging this, Schmitt and colleagues called for further investigation into potential confounding of response elaboration with item verifiability in applicant settings.

Therefore, there are three primary purposes of this study. First, we examine whether the required elaboration technique is a viable approach to control response inflation in a high-stakes selection situation where job candidates are motivated to inflate their scores on selection tests. Second, we examine whether verbal ability will moderate the relationship between elaboration and biodata item scores in a high-stakes selection situation. Third, we examine the impact of required elaboration on verifiable and nonverifiable biodata items.

Response Elaboration as a Technique to Reduce Score Inflation

Schmitt and Kunce (2002) developed RET drawing from the logic underlying accomplishment records (Hough, 1984), where applicants are required to offer evidence to prove they have job-related skills by describing past accomplishments and to provide references that can verify their accomplishments. RET builds on this by requiring job applicants to provide supporting information to justify their answers to test questions. For example, for a biodata question such as, "In the past 5 years, how many

times have you attended a conference, seminar, or workshop as a way of gaining new skills or information in an area of interest?," a required elaboration might be, "List the title(s) or provide brief description(s) of the event(s)."

Schmitt and Kunce (2002) argued that RET may decrease intentional response distortion (e.g., exaggeration, lying) and avoid unintentional response distortion (e.g., self-deception). They provided initial empirical support for the usefulness of RET by finding that requiring elaboration for biodata items reduces mean test scores by .70 to .80 standard deviations as compared to nonelaborated items. The same results were found in two replication studies by Schmitt and colleagues (Ramsay et al., 2006; Schmitt et al., 2003). For their part, Lievens and Peeters (2008) found that elaboration decreases faking on situational judgment test items with high familiarity (e.g., test takers have a prior experience or are familiar with the event or situation described in the item) by .43 standard deviations as compared to faking on nonelaborated items.

Response Elaboration and Reflection Upon Past Experiences

It is clear that response elaboration can help reduce mean test scores. Research on reflection provides an explanation as to why this is likely to occur. Reflection is a psychological process through which individuals systematically think about their experiences, make sense of such experiences, and plan for future actions (Kolb, 1984). RET requires candidates to reflect upon past experiences and is consistent with a three-staged reflection process model, which includes a return to experience, analysis, and reevaluation of the experience and learning outcomes (Boud, Keogh, & Walker, 1985). The first stage requires job applicants to return to and recall previous experiences and behaviors that are initiated by the test item and the requirement to support the endorsed answer. Autobiographical memory searches will prompt consideration of a relatively broad set of plausible supporting alternatives. Some of the recalled experiences will support the desired answer, whereas other recalled experiences will not support the desired answer.

The second stage involves critical analysis of the recalled experiences in order to identify and write down the required support of the endorsed item. Job applicants need to engage in effortful and elaborative thinking to identify the most appropriate supportive experiences from the broad set of recalled experiences. The act of writing down the supporting information makes thoughts more concrete and stable (Sedikides, Horton, & Gregg, 2007), increases accessibility of recalled experiences for subsequent analysis (Pennebaker, 2003), increases a sense of accountability (Tetlock, Skitka, & Boettger, 1989), and increases commitment of job

applicants to the content of the statements expressed (Festinger & Carlsmith, 1959).

The third stage involves a reevaluation of the endorsed answer based on the recalled prior experiences. As mentioned earlier, it is likely that job applicants recall a variety of past behaviors, including behaviors that correspond to both high and low scores on particular biodata items. Debiasing research suggests that considering alternatives or more than one point of view on oneself leads to more moderate self-evaluations (Anderson, 1982; Hirt, Kardes, & Markman, 2004). In addition, job candidates might realize at this stage that the recalled experience cannot be used to justify the inflated answer. As such, accessibility of both answer-confirming and answer-disconfirming experiences in mental form or having an insufficient number of supporting written experiences may cause applicants to realize that they cannot support the desired answer, resulting in endorsing less inflated answers on elaborated biodata items.

These stages can be described in terms of the amount of effort required to engage in the reflection processes described above (Peltier, Hay, & Drago, 2005). Three different levels of reflection have been identified and described in the literature. Descriptive reflection is the lowest level of reflection and involves looking back on experiences and describing them (Conway, 2001). Explanatory reflection involves looking back on experiences and searching for different explanations for the experiences (Ross, 1989). Critical reflection is the highest level of reflection and leads to changes in deep-ingrained beliefs and to new belief structures (Kember, McKay, Sinclair, & Wong, 2008).

Biodata items without elaboration require applicants to *descriptively* reflect upon their past experiences and behaviors and to conclude that they had those experiences or engaged in those behaviors to some extent. For example, the biodata item "How often have you looked for a new way to complete work assignments?" requires applicants to reflect upon how they typically complete work assignments and then conclude that they used new approaches to do work assignments to some extent.

Biodata items with elaboration require applicants to engage in higher level *explanatory* reflection in that they will not only descriptively reflect on their past experiences, but they will also articulate various features of their experiences and behaviors. Job applicants are thus required to generate an explanation that describes and justifies their response and then conveys these reasons (e.g., instances of past experiences) through a narrative to others. For example, a required elaboration for the biodata item, "How often have you looked for a new way to complete work assignments?" may ask job applicants to list and describe new ways they completed their work assignments. This will force applicants to retrospectively reflect on how they completed assignments in

the past, identify innovative approaches, justify to themselves that identified approaches would be considered as innovative by others, write down the identified approaches, and reevaluate whether or not the amount of identified written supporting experiences justifies the desired endorsed answer.

Thus, the descriptive reflection characterizing answers to biodata items without elaboration involves incidental attention to past experiences and is marked by holistic, automatic, and relatively fast processing at very shallow cognitive levels (Evans, 2008). The explanatory reflection characterizing answers to biodata items with elaboration requires significant attention to past experiences and behaviors and is characterized by detailed, controlled, analytic, and more time-consuming processes that demand considerable cognitive capacity (Smith & DeCoster, 2000). These two types of reflection are more likely to lead to different outcomes (Anseel, Lievens, & Schollaert, 2009). For example, effortful information processing helps people recall past experiences more accurately and self-evaluate more realistically (Tetlock, 1985; Tetlock et al., 1989). Also, Sedikides et al. (2007) found that explanatory introspection (which requires people to consider the reasons why they are the kind of person they are) reduces self-enhancement, but descriptive introspection (which requires people to describe the extent to which they do or do not possess particular traits) does not reduce self-enhancement. Critically, these effects occurred only when participants were asked to write down their thoughts. This suggests that asking candidates to elaborate their responses may reduce self-enhancement, resulting in less inflated scores. Therefore, mean biodata item responses will be lower for elaborated than nonelaborated items.

> *Hypothesis 1*: Mean biodata item responses will be lower for elaborated items than for nonelaborated versions of the same items.

The Moderating Effect of Elaboration on the Relationship Between Verbal Ability and Biodata Scores

As suggested, elaboration requires applicants to engage in explanatory reflection that involves relatively extensive and effortful information processing activity aimed at recalling and scrutinizing the information in relevant self-schema to support an endorsed answer. In addition, writing elaborations increases verbal ability demands because of the need to formulate and articulate a coherent written answer.

Applicants with high levels of verbal ability are more likely to be better at explanatory reflection because they can write in a coherent manner and

have fewer difficulties with writing about events from their past work and personal experiences. They are likely to be more well-read and have better comprehension and retention of the material, thus giving them a larger set of experiences upon which to draw. They would likely spend less time developing and writing down their supporting information, provide more detail and specificity in their elaborations, and convey information in logical and concise manner (Fowler & Kroll, 1980; Hughey, 1995; Preston & Gardner, 1967).

In addition, the increased verbal demands due to RET and explanatory reflection will make it more difficult for job applicants to inflate their responses. There is considerable evidence in the social psychology and motivated-reasoning literatures to suggest that people motivated to draw a desired conclusion or inflate their responses attempt to be rational and to construct a justification of the conclusion or the response that would be plausible to a dispassionate observer (Darley & Gross, 1983; Kunda, 1990; Pyszczynski & Greenberg, 1987). Job applicants who are motivated to enhance their answers on a biodata item with an elaboration need to consider how to justify and support the enhanced answer before providing it. However, for an applicant who enhances a biodata item with no elaboration, the consideration of "how" to justify the answer is not prominent in the decision to enhance it. Job applicants will enhance their answer on a biodata item with an elaboration only if they can offer written evidence necessary to support it. They may selectively search memory for the experiences that support their desired answer. They may creatively combine accessed knowledge or experiences or invent experiences to construct new experiences that could logically support the desired endorsed answer. Job applicants' endorsement of the enhanced answers will be constrained by their ability to construct seemingly reasonable written justifications for their inflated responses in a short period of time. In total, this suggests that the use of the RET will lead to increased verbal demands of responding, resulting in a stronger relationship between verbal ability and biodata items with elaborations.

> *Hypothesis 2*: Elaboration will moderate the relationship between verbal ability and biodata item scores such that verbal ability will be more strongly related to biodata item scores when items require elaboration.

The Effect of Item Verifiability on Responses to Biodata Items

Item verifiability can be defined as the degree to which item responses can be corroborated from an independent source, such as archival data (e.g., school transcripts, work records) or testimonies of others (e.g.,

teachers, managers, coworkers). Verifiable items include all factual, external behaviors performed in the presence of others, regardless of how difficult or unlikely it would be to obtain corroboration from witnesses (Gandy, Outerbridge, Sharf, & Dye, 1989; Mael, 1991; Stricker, 1987). When applicants answer verifiable items, they can be held accountable for their responses.

Verifiable (e.g., "What was your grade point average in college?") and nonverifiable (e.g., "How often have you looked for a new way to complete an assignment?") items provide different opportunities for job applicants to engage in response distortion. Job applicants may fake verifiable items less often for several reasons. First, faking on verifiable items represents a more severe form of deception. When applicants fake verifiable biodata items, they endorse a response option that is verifiably false. For example, endorsing a response option of college GPA equaling 4.00 is a verifiably false response for applicants whose real GPA is 3.00. There is some evidence suggesting that job applicants are less likely to engage in more severe forms of faking (e.g., lying; Becker & Colquitt, 1992; Cascio, 1975; Keating, Paterson, & Stone, 1950; Kluger & Colella, 1993; Lautenschlager, 1994; Mosel & Cozan, 1952) in favor of less severe forms of faking (e.g., exaggeration; Donovan, Dwight, & Hutz, 2003; Levashina & Campion, 2007). Second, faking on verifiable items is more detectable. The responses to verifiable items can be compared with factual data obtained from an independent source (e.g., self-reported educational degrees can be compared with the educational reports provided by the degree-issuing institution). Third, detected faking on verifiable items is more actionable. False responses on verifiable items may be exposed, and job applicants may be disqualified from the selection process. Most organizations warn job applicants that providing fraudulent information during the selection process can lead to not getting the job or being terminated if it is discovered. Moreover, faking on verifiable items and a false representation of a material fact (as opposed to an opinion or promises) is considered to be fraud by law and is illegal (Ficht & Levashina, 2008; Walsh, 2007). Fourth, it is more difficult for applicants to engage in self-deception on verifiable items because they are more objective (Paulhus & Reid, 1991), thus resulting in conscious deception (Mael, 1991). For these reasons, job applicants are less likely to engage in this more severe form of faking.

On the other hand, job applicants may fake nonverifiable items more often because the nature of these items allows applicants to engage in less severe forms of faking that are more difficult to detect and are less actionable. For instance, the biodata question: "How often do you put in extra effort to finish a group project on time?" allow applicants to engage in exaggeration and embellishment with little risk of being caught. This

is a subjective evaluation, and it is unclear how one would determine the accuracy of such a statement. In total, this suggests that job applicants will fake verifiable items less often but nonverifiable items more often, resulting in lower scores on verifiable than on nonverifiable items.

> *Hypothesis 3*: Mean biodata item responses will be lower for verifiable items than for nonverifiable items.

The Moderating Effect of Elaboration on the Impact of Item Verifiability on Biodata Scores

RET is likely to decrease score inflation via increased accountability of responses. Accountability refers to the degree to which people believe that they are responsible or can be held responsible for their actions (Sedikides, Herbst, Hardin, & Dardis, 2002). Different features of the selection tests or test items (e.g., verifiable and nonverifiable items) may be associated with higher or lower likelihood that people can be held accountable for their responses.

Social psychological research has discovered that implicitly or explicitly telling people that their judgments are open to scrutiny or evaluation by others is a key way to hold people accountable for their evaluations (McKenna & Myers, 1997; Sedikides et al., 2002). Accountable people are more likely to evaluate their work critically, anticipate counterarguments, and incorporate the most probable criticism into their evaluations (Tetlock, 1983a, 1983b), resulting in more accurate self-evaluations. RET requires applicants to provide proof of the endorsed answer by describing events or situations that support their responses. The information provided in the elaboration itself (e.g., events, names, or situations) could be subject to follow-up verification. It also suggests to job applicants that their self-evaluations are open to investigation by others who may or may not come to the same conclusion, leading to increased job applicant accountability for their responses.

Yet, because verifiable items (with or without required elaborations) can be checked and verified by others, job applicants are likely to feel more accountable for their responses. This will likely minimize the potential benefits of asking for job applicants to elaborate their responses to verifiable items, in part because they already feel highly accountable due to the nature of the item itself. That is, required elaborations are unlikely to significantly increase job applicants' accountability for their responses because such accountability is already built in to the item itself.

On the other hand, because nonverifiable items (without required elaboration) do not make it possible for responses to be verified by others, job applicants are less accountable for their responses. But asking job applicants to elaborate nonverifiable items will increase accountability

for their responses for several reasons. First, job applicants are asked to justify their answer to nonverifiable items by describing events or situations that support their responses. To respond in a less than honest manner would require the deliberate fabrication of specific events or situations. Second, the information provided in the elaboration itself (e.g., events, names, or situations) could be subject to follow-up verification, and this would enhance the salience of a dishonest response. Third, RET will indicate to job applicants that their responses to nonverifiable items are now open to investigation by others who may or may not come to the same conclusion based on the information provided in the elaboration.

> *Hypothesis 4*: Elaboration will moderate the impact of item verifiability on biodata item scores such that the use of elaboration will reduce scores on nonverifiable biodata item responses to a greater extent than those on verifiable biodata item responses.

Method

Sample

The sample consisted of 16,304 job applicants for professional entry-level U.S. government jobs. These positions entail working with the public and government officials, and members of the business community in both the United States and foreign countries, in one of several different career fields (e.g., management, trade, and public relations). Thirty-nine percent were female; 72% were White; 10% were Asian; 10% were Black; 9% were Hispanic; and 1% was of other ethnic backgrounds. Approximately 89% of the applicants had bachelor's and higher-level degrees, and a mean age of applicants was approximately 25 years. Because gender and race were not related to the biodata research item score, they were not included in the analyses as control variables.

Measures and Procedure

Biodata test.   The 88-item biodata measure was developed following a comprehensive job analysis. Nine key areas of "life experience" were identified via job analysis (e.g., interactions with others, adaptability, initiative or persistence, leadership, and so on), and biodata items were written to assess the past experiences applicants had in these domains. A typical biodata item was "When you have first moved into a new place, how much time have you spent exploring your new surroundings? (1 = *very little time*, 5 = *a great deal of time*)."

Of the 88 items, 35 required respondents to elaborate by providing written support for their responses (Schmitt & Kunce, 2002). A typical

elaborated biodata item was "In the last months, how often have you read about cultures different from your own? (1 = *never*, 5 = *very often*). If you answered 2, 3, 4, or 5, list the cultures and the materials read. List no more than four." Rational scoring was used and all biodata items were scored continuously on a 5-point scale. Internal consistency reliability of the biodata measure was .95.

To test our hypotheses, we designed a field experiment with an item level experimental manipulation. We used a 2 (elaborated/ nonelaborated) × 2 (verifiable/nonverifiable) factorial design to examine the impact of the biodata item characteristics on job applicants' responses. We developed two verifiable and two nonverifiable biodata items. The first verifiable item asked, "In the past year, how many times have you helped a new coworker or team member get adjusted to the job or team?," with response options ranging from 1 = 0 to 5 = *4 or more times*. The second verifiable item asked, "How often have you been in a leadership role in organizations to which you have belonged?," with response options ranging from 1 = *never* to 5 = *very often*. The first nonverifiable item asked, "How often do others rely on your guidance during stressful situations?" and the second nonverifiable item asked, "In situations where you have a disagreement with a friend or coworker, how often do you choose to talk about the issue, instead of ignoring it or working separately?" Response options ranged from 1 = *never* to 5 = *very often* for both nonverifiable items. These four verifiable and nonverifiable items had elaborated and nonelaborated versions, for a total of eight different experimental items. Each applicant randomly received one of the eight experimental items. Thus, all applicants responded to a single experimental item that was either elaborated or nonelaborated and verifiable or nonverifiable. Because items were randomly assigned to applicants, difference in item responses can be attributed to the item-level manipulation (Bertrand & Mullainathan, 2004; Montgomery, 1997). Thus, each applicant answered 53 nonelaborated items, 35 elaborated items, and one experimental biodata item that was either elaborated or nonelaborated and verifiable or nonverifiable. Job applicants were not aware that some of the items in the biodata instrument were being used for research purposes.

Finally, applicants were given a warning that their responses could be verified and that any attempts to falsify information would be used as a basis for not employing them or dismissing them after they have begun work. Such warnings are commonly used in operational selection contexts.

Verbal ability.   The verbal ability test consisted of 65 multiple-choice items that measured word usage, vocabulary, verbal reasoning, and writing style. All items had four options. Internal consistency reliability was .92, and scores were standardized (mean = 50, *SD* = 10).

TABLE 1

Means, Standard Deviations, and Effect Sizes of the Impact of Item Elaboration and Item Verifiability on Biodata Research Item Scores

Factor	*M*	*SD*	*d*	95% CI for *d*	*F*
Verifiability (V)					753.10
Nonverifiable (*N* = 8,215)	3.97	.89			
Verifiable	3.50	1.28	.43	(.40; .46)	
Elaboration (E)					108.29
Nonelaborated (*N* = 8,076)	3.83	1.10			
Elaborated	3.65	1.14	.16	(.13; .19)	
Interaction V × E					11.59
Verifiable					
Nonelaborated (*N* = 4,034)	3.56	1.26			
Elaborated	3.44	1.29	.09	(.05; .14)	
Nonverifiable					
Nonelaborated (*N* = 4,042)	4.09	.83			
Elaborated	3.85	.93	.27	(.23; .32)	

Note. All means are statistically different, $p < .001$. *d* represents the standardized mean difference between nonverifiable and verifiable, and nonelaborated and elaborated items. All *F* statistics are statistically significant, $p < .001$.
$N = 16,304$.

Results

    To test the hypotheses of this study, we performed analysis of variance (ANOVA) with the experimental biodata item score as the dependent variable and item verifiability and required elaboration as independent variables. The 2 (verifiable/nonverifiable) × 2 (elaborated/nonelaborated) ANOVA revealed a significant verifiability effect, $F (1, 16303) = 753.10$, $p < .0001$; a significant elaboration effect, $F (1, 16303) = 108.29$, $p < .0001$; and a significant interaction of Verifiability × Elaboration, $F (1, 16303) = 11.59, p < .001$.

    Hypothesis 1 suggested that mean biodata item responses will be lower for elaborated items than for nonelaborated versions of the same items. Table 1 contains the means and standard deviations across different biodata item types. We also estimated the 95% confidence intervals for the *d* values (Algina, Keselman, & Penfield, 2006; Cumming & Finch, 2001). As can be seen, requiring job applicants to elaborate their answers to items produced lower item means compared with the means of the same items with no elaboration ($d = .16$; 95% CI(.13; .19), $p < .0001$), providing support for Hypothesis 1.

    Hypothesis 2 suggested that elaboration will moderate the relationship between verbal ability and biodata items score. To test this hypothesis, we conducted hierarchical regression analysis with biodata research item

TABLE 2
Interaction of Item Elaboration and Verbal Ability

	B	SE B	T
Step 1			
Verifiability	−.47	.02	−27.34**
Elaboration	−.18	.02	−10.46**
Verbal ability	.004	.001	4.40**
Step 2			
Verifiability	−.47	.02	−27.36**
Elaboration	−.18	.02	−10.47**
Verbal ability	.002	.001	1.66
Elaboration × Verbal ability	.004	.002	2.06*

Note. $R^2 = 0.0510$, adjusted $R^2 = 0.0508$ for Step 1; $\Delta R^2 = 0.0003^*$, ΔR^2 (adjusted) = 0.0002^* for Step 2.
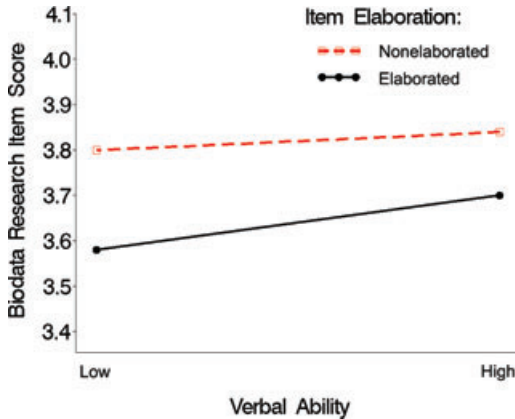$^* p < 0.05.$ $^{**} p < 0.001.$

Figure 1: **Moderating Effect of Item Elaboration on the Relationship Between Verbal Ability and Biodata Research Item Scores.**

scores regressed on verifiability, elaboration, and verbal ability scores in Step 1 and the elaboration by verbal ability interaction in Step 2. Verbal ability was centered prior to conducting regression analyses to remove nonessential multicollinearity (Cohen, Cohen, West, & Aiken, 2003). Table 2 shows that the interaction was significant ($\beta = .004$, $p < .05$). Following the procedures suggested by Cohen and Cohen (1983), we plotted the verbal ability by elaboration interaction term for verbal ability scores at 1 *SD* below and above their mean. As illustrated in Figure 1, the slope of the solid line indicates a positive relationship between verbal ability and biodata research items with elaboration. We further probed the
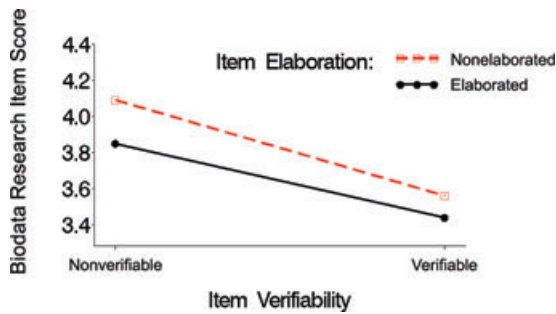
Figure 2: **Moderating Effect of Item Elaboration on the Impact of Item Verifiability on Biodata Research Item Scores.**

nature of the interaction by conducting a simple slope analysis (Cohen et al., 2003). Supporting Hypothesis 2, the relationship between verbal ability scores and biodata research item scores was positive for job applicants who answered research biodata items with elaboration ($\beta = .006$, $t(16303) = 4.56$, $p < .0001$). This relationship was positive but not significant for applicants who answered research biodata items without required elaboration ($\beta = .002$, $t(16303) = 1.66$, $p = .10$). These findings support Hypothesis 2.

Hypothesis 3 suggested that mean biodata item responses will be lower for verifiable items than for nonverifiable items. The difference between nonverifiable and verifiable item means was 0.43 standard deviations (95% CI (.40; .46), $p < .001$, Table 2), supporting Hypothesis 3. Hypothesis 4 suggested that required elaboration will reduce means of nonverifiable items to a greater extent than those of verifiable items. As can be seen from Table 1 and Figure 2, requiring job applicants to elaborate their answers to nonverifiable items significantly reduced item means ($d = .27$, $p < .001$, 95% CI (.23; .32)). Although requiring job applicants to elaborate on their answers to verifiable items also significantly reduced item means, the effect size of elaboration on these types of items is significantly lower ($d = 0.09$, $p < .001$, 95% CI (.05; .14)), providing full support for Hypothesis 4.

Discussion

Summary of Findings

In this research, we sought to explore the underlying mechanisms through which RET influences applicant responding in an actual selection situation. We proposed that RET requires applicants to engage in explanatory reflection and generate a written explanation that describes and justifies their responses.

Consistent with Hypothesis 1, we found that elaboration reduces bio-data item mean scores in high-stakes selection context. Our effect sizes ($d = .16$), however, were lower than those found in past experimental research ($d = .70$ to $d = .80$ in Schmitt & Kunce, 2002; $d = .80$ in Schmitt et al., 2003; $d = .43$ in Lievens & Peeters, 2008). There are several plausible explanations for this.

First, past research demonstrates that although applicants are highly motivated to fake their responses, the degree to which applicants fake in the high-stakes selection context is less than the degree to which research participants distort their scores when instructed to fake (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006). Compared with research participants, job applicants selectively inflate their answers to certain types of items in a measure. For example, they may inflate their scores on items that they perceive to be job-related or could be used to create the best job applicant image in a way that minimizes faking detection, thus appearing similar to the scores for candidates who honestly describe their credentials (Vasilopoulos et al., 2005). As such, the effect sizes of real-life faking are substantially smaller than those of induced faking, where effect sizes in applicant settings typically are a half to one standard deviation lower (Rosse et al., 1998; Viswesvaran & Ones, 1999).

Second, the use of a warning may have reduced the effect sizes. Research demonstrates that a warning of response verification leads to de-flated scores on noncognitive measures (Vasilopoulos et al., 2005). In this study, job applicants were warned that their responses on a biodata measure may be verified and that any attempts to falsify information would be used as a basis for not employing them or dismissing them after they have begun work. This warning may have reduced levels of faking and thus weakened the RET effect sizes. We used warnings, however, because they are commonly used in testing settings, so our procedures are consistent with past selection practice (enhancing the external validity of our study). Third, we examined the impact of RET on job applicants' responses in a very large sample of real applicants, which allowed us to yield reliable results and to control for statistical artifacts (e.g., sampling error) that might be present in past studies on elaboration (Murphy, 2000; Schmidt & Hunter, 2003), potentially attenuating the effect sizes.

Consistent with Hypothesis 2, we found that elaboration moderates the relationship between verbal ability and biodata scores. Although the moderation was statistically significant, the effect size of this moderation was small. This finding suggests that requiring applicants to elaborate their responses does not result in a strong relationship between biodata and verbal ability scores and, thus, will not inadvertently change the constructs assessed with biodata measure. Consistent with Hypothesis 3, we found that job applicants score significantly higher on nonverifiable items

than verifiable items, and the difference is about .43 standard deviation units. This finding is consistent with past research indicating that job applicants are less likely to inflate their responses on verifiable items and more likely to inflate their responses on nonverifiable items (Becker & Colquitt, 1992). Finally, consistent with Hypothesis 4, we found that the effect of required elaboration is greater for nonverifiable items ($d = .27$) than for verifiable items ($d = .09$). This study addresses a potential study confound often present in studies on elaboration. Across existing studies of elaboration, the use of RET has been confounded with item verifiability. The results of this study indicate that RET decreases score inflation on nonverifiable items more than on verifiable items, which we suggest is due to increased accountability. Verifiable items make job applicants accountable for their responses. Job applicants may think that it is "unsafe" to inflate responses on verifiable biodata items because it is possible that the potential employer will detect the response inflation and take action (e.g., remove them from the selection process). Nonverifiable items are more prone to response inflation because they cannot hold job applicants accountable for their responses. Because there are some constructs (e.g., leadership, initiative) that are measured primarily by nonverifiable items (Harold, McFarland, & Weekley, 2006), asking job candidates to elaborate their responses may be a viable strategy to increase job applicants' accountability and decrease score inflation.

Implications for Theory and Practice

There are several theoretical and practical implications of this research. First, we directly investigated RET in a high-stakes selection context for highly desirable jobs. Only four empirical studies (Lievens & Peeters, 2008; Ramsay et al., 2006; Schmitt & Kunce, 2002; Schmitt et al., 2003) have been conducted to assess the usefulness of RET. Importantly, none of this past research has been conducted with actual job applicants. Our results support the use of RET to reduce intentional distortion on biodata measures in high-stakes settings. In addition, the effect of required elaboration for nonverifiable items ($d = .27$) is three times greater than for verifiable items, suggesting that the RET technique will have greater utility when it is used primarily for nonverifiable items in practice.

Second, we examined the impact of RET at the item level and not the scale level (the handful of published RET studies have focused on the scale level) because applicants "respond to (and fake) individual items, not scales" (Zickar & Robie, 1999, p. 552). Furthermore, the item-level information should be useful to practitioners involved in the development and assessment of biodata items or researchers involved in the development of faking resistant selection measures.

Third, although statistically significant, we found somewhat small effect sizes, especially for verbal ability. Accordingly, one must be careful not to overinterpret these findings. Yet, these small effects may still have some practical and theoretical implications (Prentice & Miller, 1992). Particularly, the size of the effect can affect the practical implications that can be drawn based on our results. One of the appealing characteristics of biodata measures is that they do not have adverse impact. Our results indicate that RET may significantly increase the verbal demands of responding. However, because the effect size of verbal ability is very small, it is unlikely that using biodata items with elaboration will result in elevated levels of adverse impact.

Limitations and Future Research

There are several issues in this research that future research should address. First, one limitation of our study is that we did not have any criterion data in order to examine whether the use of RET enhances the criterion-related validity of biodata measures. Although Schmitt et al. (2003) found that RET did not have a practically significant effect on criterion-related validity of a biodata measure in a student sample, the effects of RET on validity are unknown in a high-stakes selection situation.

Second, there are possible alternative explanations for the RET-induced score reduction. For example, job applicants may simply have experienced cognitive fatigue or had difficulties in recalling supporting information. Yet, past research demonstrated that increased cognitive fatigue does not necessarily lead to decreased performance (Myers, 1937; Thorndike, 1900). Moreover, some have reported increased performance as time on task, subjective fatigue, or test length increases (Ackerman & Kanfer, 2009; Davis, 1946). Finally, Ramsay et al. (2006) found that respondent ability to recall specific events did not explain the lower scores on elaborated biodata items in a biodata measure that included both elaborated and nonelaborated items. Yet, because we are unable to rule out these alternative explanations, future research should explore them to determine if they are in fact operating.

Third, one of the reviewers noted that item verifiability in this study might be partially confounded with the base rate of behavior as reflected in the response options for the respective items. Because one of the two verifiable items ("In the past year, how many times have you helped a new coworker or team member get adjusted to the job or team?") constrains the response options to a restricted time frame (thus constraining the reporting of frequency of behaviors) and both nonverifiable items were not commensurately constrained, we might expect lower mean scores for the verifiable items than for the nonverifiable items. Thus, the support for

Hypothesis 3 is tentative and should be replicated by future studies that control for item and response characteristics.

Fourth, although the verifiable items used in our study were not easily verified, they do represent items that could theoretically be verified. In other words, it might be difficult to verify the responses, but it could be done. Of course, one limitation associated with using these kinds of items is that it represents a somewhat weaker manipulation of verifiability, as some applicants may have recognized the difficulty associated with verifying responses. Thus, it may not have reduced response distortion as much as more clearly verifiable items. As noted by others, item verifiability represents a continuum from nonverifiability to clear verifiability (Mael, 1991). This suggests that our research represents a conservative test of the effects of item verifiability. Future research would be advised to explore how variation in item verifiability affects applicant responding.

Fifth, future research should investigate what types of response distortion are mitigated by RET. Researchers have identified two types of response distortion: self-deception or unintentional distortion of responses and impression management or intentional response distortion (Paulhus, 1984; Zerbe & Paulhus, 1987). RET could reduce both intentional and unintentional tendencies to distort self-reports. Intentional response distortion may be decreased because required elaboration makes job applicants more accountable for their responses. Unintentional response distortion may be decreased because elaboration is likely to enhance controlled information processing. When people expect that they need to justify their judgments to others, they become more vigilant information processors (McKenna & Myers, 1997) and engage in deeper and more careful information processing (Cvetkovich, 1978; Tetlock, 1985). Elaborations require applicants to recall and describe concrete and contextualized examples and events to justify their responses. When applicants answer biodata items with elaborations, they are more likely to think about questions in a more detailed and complex manner (e.g., consider more alternatives, evaluate their experiences from more perspectives) as opposed to a more abstract, simple, and idealized manner when they answer biodata items with no elaborations, leading to decreased self-deception. As such, RET should make people with high levels of self-deception recall past experiences more accurately and self-evaluate more realistically.

Sixth, future research should investigate whether RET decreases score inflation across different types of biodata items (Mael, 1991). Different biodata items (e.g., controllable vs. noncontrollable) may provide different opportunities to make job applicants accountable for their responses. For example, asking applicants to elaborate controllable items (items that measure behaviors that applicants chose to perform or not to perform) can make applicants accountable for their responses and lead to decreased

scores. On the other hand, asking job applicants to elaborate noncontrollable items (items that measure actions that happened to or were done to the applicant) cannot make applicants accountable for their responses because they may perceive that the response to such items is significantly influenced by factors external to themselves.

Seventh, biodata measures can potentially measure a host of different underlying constructs, including cognitive ability, knowledge, motivation, personality, values, and interests (Brown & Campion, 1994; Mael & Ashforth, 1995). These different constructs may vary with respect to the extent to which they provide applicants with opportunities to engage in response distortion. Biodata items that measure cognitive constructs (e.g., "How many languages can you speak at a conversational level or better?") are less likely to be faked. These items have objectively correct answers, and thus, they are more verifiable. Biodata items that measure noncognitive constructs (e.g., "How often do you put in extra time on a project?"), on the other hand, are more likely to be faked. Because these items do not have objectively correct answers, they are less verifiable. Thus, RET may reduce scores on biodata items that measure noncognitive constructs to a greater extent than scores on biodata items that measure cognitive constructs. This is an area that would benefit from additional research.

Finally, future research should investigate job candidate reactions to the use of RET. It is possible that some job applicants may have a negative reaction to this technique because they are required to exert extra effort to recall situations and events in order to elaborate their responses. On the other hand, some applicants may have positive reactions to this technique because they are given an opportunity to justify their responses and to demonstrate more adequately their knowledge, skills, and abilities in the testing situation (Lievens & Peeters, 2008; Schleicher, Venkataramani, Morgeson, & Campion, 2006) and keep other candidates more honest. These are possibilities worth exploring.

REFERENCES

Ackerman PL, Kanfer R. (2009). Test length and cognitive fatigue: An empirical examination of effects of performance and test-taker reactions. *Journal of Experimental Psychology: Applied*, *15*, 163–181.

Algina J, Keselman HJ, Penfield RD. (2006). Confidence interval coverage for Cohen's effect size statistics. *Educational and Psychological Measurement*, *66*, 945–960.

Alliger GM, Dwight SA. (2000). A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. *Educational and Psychological Measurement*, *60*, 59–72.

Anderson CA. (1982). Inoculation and counterexplanation: Debiasing techniques in the perseverance of social theories. *Social Cognition*, *1*, 126–139.

Anderson CD, Warner JL, Spencer CC. (1984). Inflation bias in self-assessment examinations: Implications for valid employee selection. *Journal of Applied Psychology*, *69*, 574–580.

Anseel F, Lievens F, Schollaert E. (2009). Reflection as a strategy to enhance task performance after feedback. *Organizational Behavior and Human Decision Processes*, *110*, 23–35.

Becker TE, Colquitt AL. (1992). Potential versus actual faking of a biodata form: An analysis along several dimensions of item type. PERSONNEL PSYCHOLOGY, *45*, 389–406.

Bertrand M, Mullainathan S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review*, *94*, 991–1013.

Birkeland SA, Manson TM, Kisamore JL, Brannick MT, Smith MA. (2006). A meta-analytic investigation of applicant faking on personality measures. *International Journal of Selection and Assessment*, *14*, 317–335.

Bobko P, Roth PL, Potosky D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. PERSONNEL PSYCHOLOGY, *52*, 561–590.

Boud D, Keogh R, Walker D. (1985). *Reflection: Turning experience into learning*. London, UK: Kogan Page.

Brown BK, Campion MA. (1994). Biodata phenomenology: Recruiters' perceptions and use of biographical information in resume screening. *Journal of Applied Psychology*, *76*, 897–908.

Cascio WF. (1975). Accuracy of verifiable biographical information blank responses. *Journal of Applied Psychology*, *60*, 767–769.

Christiansen ND, Burns GN, Montgomery GE. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, *18*, 267–307.

Cohen J, Cohen P. (1983). *Applied multiple regression/ correlation analysis for the behavioral sciences* (3rd ed.). Hillsdale, NJ: Erlbaum.

Cohen J, Cohen P, West SG, Aiken LS. (2003). *Applied multiple regression/ correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Conway PE. (2001). Anticipatory reflection while learning to teach: From a temporally truncated to a temporally distributed model of reflection in teaching education. *Teaching and Teacher Education*, *17*, 89–106.

Cumming G, Finch S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement*, *61*, 532–574.

Crowne D, Marlowe D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349–354.

Cvetkovich G. (1978). Cognitive accommodation, language, and social responsibility. *Social Psychology*, *41*, 149–155.

Darley JM, Gross PH. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, *44*, 20–33.

Davis DR. (1946). The disorganization of behavior in fatigue. *Journal of Neurology, Neurosurgery, and Psychiatry*, *9*, 23–29.

Donovan JJ, Dwight SA, Hutz GM. (2003). An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance*, *16*, 81–106.

Dwight SA, Donovan JJ. (2003). Do warnings not to fake reduce faking? *Human Performance*, *16*, 1–23.

Ellingson JE, Sackett PR, Hough LM. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, *84*, 155–166.

Evans J. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278.

Festinger A, Carlsmith J. (1959). Cognitive consequences of forced compliance. *Journal of Personality and Social Psychology*, *58*, 203–210.

Ficht LS, Levashina J. (2008). *When lying, cheating, and stealing isn't necessarily illegal: The need to adopt a commercial misrepresentation standard in employment cases*. Paper presented at the Annual Conference of the Academy of Legal Studies in Business, Long Beach, CA.

Fowler B, Kroll BM. (1980). Relationship of apprehension about writing to performance as measured by grades in a college course on composition. *Psychological Reports*, *46*, 583–586.

Gandy JA, Outerbridge AN, Sharf JG, Dye DA. (1989). *Development and initial validation of the Individual Achievement Record*. Washington, DC: U.S. Office of Personnel Management.

Harold CM, McFarland LA, Weekley JA. (2006). The validity of verifiable and non-verifiable biodata items: An examination across applicants and incumbents. *International Journal of Selection and Assessment*, *14*, 336–346.

Hirt ER, Kardes FR, Markman KD. (2004). Activating a mental simulation mind-set through generation of alternatives: Implications for debiasing in related and unrelated domains. *Journal of Experimental Social Psychology*, *40*, 374–383.

Hough LM. (1984). Development and evaluation of the "accomplishment record" method of selecting and promoting professionals. *Journal of Applied Psychology*, 69, 135–146.

Hughey JB. (1995). The effects of facilitative teacher response and revision strategies on adolescent writing achievement. *Dissertation Abstracts International, Section A: Humanities and Social Sciences*, *55*, 2325.

Hunter JE, Hunter RF. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72–98.

Keating E, Paterson DC, Stone HC. (1950). Validity of work histories obtained by interview. *Journal of Applied Psychology*, *34*, 6–11.

Kember D, McKay J, Sinclair K, Wong FKY. (2008). A four-category scheme for coding and assessing the level of reflection in written work. *Assessment & Evaluation in Higher Education*, *33*, 369–379.

Kluger AN, Colella A. (1993). Beyond the mean bias: The effect of warning against faking on biodata item variances. PERSONNEL PSYCHOLOGY, *46*, 763–780.

Kolb DA. (1984) *Experiential learning*. Englewood Cliffs, NJ: Prentice Hall.

Kunda Z. (1990). The case of motivated reasoning. *Psychological Bulletin*, *108*, 480–498.

Lautenschlager GJ. (1994). Accuracy and faking of background data. In Stokes GA, Mumford MD, Owens WA (Eds.), *Biodata handbook* (pp. 391–419). Palo Alto, CA: Consulting Psychologists Press.

Levashina J, Campion MA. (2007). Measuring faking in the employment interview: Development and validation of an interview faking behavior scale. *Journal of Applied Psychology*, *92*, 1638–1656.

Levashina J, Morgeson FP, Campion MA. (2009). They don't do it often, but they do it well: Exploring the relationship between applicant mental abilities and faking. *International Journal of Selection and Assessment, 17*, 271–281.

Lievens F, Peeters H. (2008). Impact of elaboration on responding to situational judgment test items. *International Journal of Selection and Assessment*, *16*, 345–355.

Mael FA. (1991). A conceptual rationale for the domain and attributes of biodata items. PERSONNEL PSYCHOLOGY, *44*, 763–792.

Mael FA, Ashforth BE. (1995). Loyal from day one: Biodata, organizational identification, and turnover among newcomers. PERSONNEL PSYCHOLOGY, *48*, 309–333.

McFarland LA, Ryan AM. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, *85*, 812–821.

McKenna FP, Myers LB. (1997). Illusory self-assessment-Can they be reduced? *British Journal of Psychology*, *88*, 39–51.

Montgomery DC. (1997). *Design and analysis of experiments* (4th ed.). New York, NY: John Wiley & Sons.

Morgeson FP, Campion MA, Dipboye RL, Hollenbeck JR, Murphy K, Schmitt N. (2007). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. PERSONNEL PSYCHOLOGY, *60*, 1029–1049.

Mosel JN, Cozan LW. (1952). The accuracy of application blank work histories. *Journal of Applied Psychology*, *36*, 365–369.

Murphy K. (2000). Impact of assessments of validity generalization and situational specificity on the science and practice of personnel selection. *International Journal of Selection and Assessment*, *8*, 194–206.

Myers CS. (1937). Conceptions of mental fatigue. *American Journal of Psychology*, *50*, 296–306.

Paulhus DL. (1984). Two-component models of socially desirable responding. *Journal of Personality & Social Psychology*, *46*, 598–609.

Paulhus DL. (1998). *Paulhus deception scales PDS: The balanced inventory of desirable responding-7 manual.* Toronto, ON: Multi-Health Systems Inc.

Paulhus DL, Reid DB. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, *60*, 307–317.

Peeters H, Lievens F. (2005). Situational judgment tests and their predictiveness of college students' success: The influence of faking. *Educational and Psychological Measurement*, *65*, 70–89.

Peltier JW, Hay A, Drago W. (2005). The reflection learning continuum: Reflecting on reflection. *Journal of Marketing Education*, *27*, 250–263.

Pennebaker J. (2003). Writing about emotional experiences as a therapeutic process. In Salovey P, Rothman JA (Eds.), *Social psychology of health: Key readings in social psychology* (pp. 362–368). New York, NY: Psychology Press.

Prentice DA, Miller DT. (1992). When small effects are impressive. *Psychological Bulletin*, *112*, 160–164.

Preston JM, Gardner RC. (1967). Dimensions of oral and written language fluency. *Journal of Verbal Learning & Verbal Behavior*, *6*, 936–945.

Pyszczynski T, Greenberg J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. In Berkowitz L (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 297–340). New York, NY: Academic Press.

Ramsay LJ, Schmitt N, Oswald FL, Kim BH, Gillespie MA. (2006). The impact of situational context variables on responses to biodata and situational judgment inventory items. *Psychology Science*, *48*, 268–287.

Reilly RR, Chao GT. (1982). Validity and fairness of some alternative employee selection procedures. PERSONNEL PSYCHOLOGY, *35*, 1–63.

Ross DD. (1989). First step in developing a reflective approach. *Journal of Teacher Education*, *40*, 22–30.

Rosse JG, Stecher MD, Miller JL, Levin RA. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, *83*, 634–644.

Roth PL, Bevier CA, Bobko P, Switzer FS, III, Tyler P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analyses. PERSONNEL PSYCHOLOGY, *54*, 297–330.

Schleicher DJ, Venkataramani V, Morgeson FP, Campion MA. (2006). So you didn't get the job…Now what do you think? Examining opportunity-to-perform fairness perceptions. PERSONNEL PSYCHOLOGY, *59*, 559–590.

Schmidt FL. (1988). The problem of group differences in ability test scores in employment selection. *Journal of Vocational Behavior*, *33*, 272–292.

Schmidt FL, Hunter JE. (2003). History, development, evolution, and impact of validity generalization and meta-analysis methods, 1975–2002. In Murphy KR (Ed.), *Validity generalization: A critical review* (pp. 31–66). Hillsdale, NJ: Erlbaum.

Schmitt N, Kunce C. (2002). The effects of required elaboration of answers to biodata questions. PERSONNEL PSYCHOLOGY, *55*, 569–587.

Schmitt N, Oswald FL. (2006). The impact of corrections for faking on the validity of noncognitive measures in selection settings. *Journal of Applied Psychology*, *91*, 613–621.

Schmitt N, Oswald FL, Kim BH, Gillespie MA, Ramsay LJ, Yoo TY. (2003). Impact of elaboration on social desirable responding and the validity of biodata measures. *Journal of Applied Psychology*, *88*, 979–988.

Schmitt N, Rogers W, Chan D, Sheppard L, Jennings D. (1997). Adverse impact and predictive efficiency using various predictor combinations. *Journal of Applied Psychology*, *82*, 719–730.

Sedikides C, Herbst KC, Hardin DP, Dardis GJ. (2002). Accountability as a deterrent to self-enhancement: The search for mechanisms. *Journal of Personality and Social Psychology*, *83*, 592–605.

Sedikides C, Horton RS, Gregg AP. (2007). The why's the limit: Curtailing self-enhancement with explanatory introspection. *Journal of Personality*, *75*, 783–824.

Seeman W. (1952). "Subtlety" in structured personality tests. *Journal of Consulting Psychology*, *16*, 278–283.

Smith ER, DeCoster J. (2000). Dual-process model in social and cognitive psychology: Conceptual integration and links to underlying memory system. *Personality and Social Psychology Review*, *4*, 108–131.

Stark S, Chernyshenko OS, Chan K, Lee WC, Drasgow F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology*, *86*, 943–953.

Stricker LJ. (1987, November). *Developing a biographical measure to assess leadership potential*. Paper presented at the Annual Meeting of the Military Testing Association, Ottawa, Ontario.

Tetlock PE. (1983a). Accountability and complexity of thought. *Journal of Personality and Social Psychology*, *45*, 74–83.

Tetlock PE. (1983b). Accountability and the perseverance of first impressions. *Social Psychology Quarterly*, *46*, 285–292.

Tetlock PE. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly*, *48*, 227–236.

Tetlock PE, Skitka L, Boettger R. (1989). Social and cognitive strategies for coping with accountability: Conformity, complexity and bolstering. *Journal of Personality and Social Psychology*, *57*, 632–640.

Thorndike EL. (1900). Mental fatigue. I. *Psychological Review*, *7*, 547–579.

Travers RMW. (1951). A critical review of the forced-choice technique. *Psychological Bulletin*, *48*, 62–70.

Vasilopoulos NL, Cucina JM, McElreath JM. (2005). Do warnings of response verification moderate the relationship between personality and cognitive ability? *Journal of Applied Psychology*, *90*, 306–322.

Vasilopoulos NL, Cucina JM, Dyomina NV, Morewitz CL. (2006). Forced-choice personality tests: A measure of personality and cognitive ability? *Human Performance*, *19*, 175–199.

Viswesvaran C, Ones DS. (1999). Meta-analysis of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, *59*, 197–210.

Walsh DJ. (2007). *Employment law for human resource practice* (2nd ed.). Eagan, MN: Thomson-West.

Zerbe WJ, Paulhus DL. (1987). Socially desirable responding in organizational behavior: A reconception. *Academy of Management Review*, *12*, 250–264.

Zickar MJ, Robie C. (1999). Modeling faking good on personality items: An item level analysis. *Journal of Applied Psychology*, *84*, 551–563.